EXPLORING THE POTENTIAL OF OPEN-DATA FOR OCEANS MONITORING WITH AI ANALYTICS

L. W. Mdakane[1,*], B. Sibolla[1], S Haupt[1]

[1]CSIR, NextGen Ent and Institutions - Spatial Information Systems Group,
Pretoria, South Africa

KEY WORDS: Vessel Tracking, Remote Sensing, Satellite Open-Data, AI Analytics, Cloud-Based Geoprocessing, Ocean Monitoring.

ABSTRACT:

Monitoring vessel activity in the ocean plays an essential role in ensuring maritime safety, environmental monitoring, and fisheries management. This article explores the importance of vessel tracking and investigates the potential of using open data for ocean monitoring using AI analysis. As satellite data demand increases, new technologies are being developed to address challenges and improve accessibility for a wider user base. This article discusses the use of AI analytics to process multiple datasets, identify ships, and track their movements. In addition, it emphasises the importance of cloud geoprocessing for accessing and analysing vast spatial data, resulting in improved decision-making and operational efficiency. In general, this article provides information on how vessel tracking can be improved for maritime safety, security, and environmental protection using open remote sensing and AI-based analysis datasets.

1. INTRODUCTION

The South African Exclusive Economic Zone (EEZ) covers an area of 1.5 million square kilometres and is an important part of the country's economic resources and trade routes (Struwig et al., 2023). Effective surveillance and protection of South Africa's coastal waters and EEZ requires advanced knowledge, forecasting capabilities, and the ability to assess environmental impacts, such as oil spills, harmful blooms, and coastal erosion, which can have severe consequences for human health, economic activities, and the environment (Vreÿ et al., 2021).

The tracking of ships is a critical component of maritime transportation management and plays an essential role in improving the safety, efficiency, and sustainability of maritime operations (Kristiansen, 2013). It allows authorities to monitor vessel movements and respond to emergencies, such as accidents or piracy incidents, in a timely manner. However, there are challenges when tracking vessels such as those related to illegal fishing or pollution. Open data and artificial intelligence can play a crucial role in improving vessel tracking and protecting our oceans (Brett et al., 2020, Katija et al., 2022, Nordling, 2017).


NATURE | NEWS

South Africa tackles crime at sea with ship-spotting satellites

Automated vessel-tracking system aims to spy poachers and smugglers.

Linda Nordling

10 November 2017

Rights & Permissions

Figure 1: A success story of using satellite data to tackle crime at sea. Source (Nordling, 2017)

1.1 AIS and Open-Data for Ship Detection

The use of Automatic Identification System (AIS) data has expanded beyond studies related to navigation, including trade flow estimates, emission tracking, and vessel performance monitoring. Satellite AIS data offers extensive coverage of the global commercial fleet, providing real-time positioning and identification information for ships. AIS enhancements can be achieved through the integration of additional data sources and advanced analysis techniques, driving the digital transformation of the shipping industry (Yang et al., 2019).

Although the high cost and limited availability of satellite data can present challenges, there are multiple open data sources that have proven successful in tracking vessel activity in the oceans (Bo et al., 2021, Hoeser et al., 2020). Figure (1) shows an example of a successful story of tracking illegal vessel activity using satellite data (Nordling, 2017). These open data sources include satellite imagery from providers such as ESA Sentinels, Planet, Digital Globe, and NOAA. These satellites capture high-resolution optical imagery or synthetic aperture radar (SAR) data, enabling the detection and tracking of vessels based on their visual characteristics or radar reflections. By leveraging open satellite data, researchers and practitioners can overcome the cost and accessibility barriers associated with proprietary datasets, facilitating a wider range of applications in maritime monitoring, safety, and management (Bhat and Huang, 2021, Li et al., 2022).

*email: lmdakane@csir.co.za

1.2 Open-data and AI Analytic for Vessel Tracking

Monitoring and analysing vessel activity in the vast oceans poses significant challenges due to their immense size and the high number of vessels traversing them. However, the use of AI analytics enables the extraction of valuable information from the locations and movements of vessels, facilitating improved situational awareness, risk assessment, and decision-making processes.

The effectiveness of AI analytics in ship identification has been demonstrated through the successful implementation of multiple machine learning (ML) classifiers, including support vector machine (SVM), random forest classifier (RFC), linear discriminant analysis (LDA), logistic regression (LR), K-nearest neighbours (KNN) and Gaussian Naive Bayes-based classifiers. These ML classifiers employ features derived from the histogram of orientated gradients (HOG) to achieve accurate identification and classification of ships (Wang et al., 2021, Li et al., 2022).

1.3 Cloud-based Geoprocessing

Cloud-based geoprocessing platforms offer significant advantages in terms of data accessibility and processing capabilities (Huang, 2020). Researchers no longer need to download and manage large data sets, as data search and browsing can be easily performed in cloud-based archives. These platforms enable the processing and analysis of large volumes of data, providing valuable insight into various aspects of the ocean (Ndehedehe, 2022). With the availability of cloud solutions, such as Google Earth Engine, NASA Earth Exchange, or ESA Cloud Toolbox, accessing and processing large datasets has become easier for a wider range of users. By eliminating barriers associated with data management, these platforms facilitate efficient access to data (Klein et al., 2017).

1.4 Research Objectives

The tracking of the activity of vessels in our oceans plays a crucial role in various domains, including maritime safety, environmental monitoring, and fisheries management. In this paper, we explore the potential of open data combined with cloud-based geoprocessing and AI analytics to track vessel activity in our oceans, exploring the benefits and challenges associated with this approach, and highlighting its implications for maritime monitoring and management.

2. STUDY AREA

South Africa's extensive coastline, which touches the Indian Ocean to the east and the South Atlantic to the west, provides three-way access for cargo and cruise ships of all sizes. Along this 1,739-mile stretch, several major ports have emerged as crucial drivers of the country's economy. Four of these ports were selected for the study, based on their vital role in facilitating trade, supporting industries and contributing to overall national development, see Figure (2).

- Port of Durban: Located on the eastern coast of South Africa, the Port of Durban has the highest vessel traffic in Africa and is the largest port in the country.
- Port of Cape Town: Located in the southwest region of South Africa, Cape Town is a major port and an important container and cargo hub. It serves as a key point for both regional and international trade.

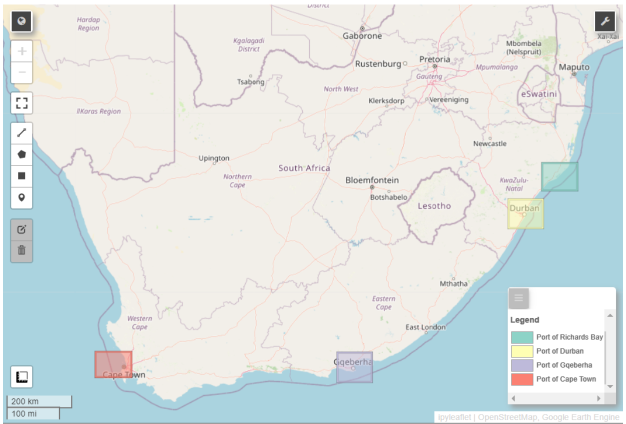

Figure 2: The study areas of interest for the selected South Africa Ports.

- Port of Richards Bay: Located on the northeast coast of South Africa, Richards Bay Port is a specialised port primarily dedicated to the export of bulk commodities. It is one of the largest coal export terminals in the world and facilitates the shipment of significant volumes of coal to various destinations throughout the world.
- Port Elizabeth (Gqeberha): Located in the Eastern Cape province of South Africa, Port Elizabeth is a major seaport that serves as a gateway to trade in the region. It offers a variety of facilities and services for various types of cargo, including containers, dry bulk, and liquid bulk shipments.

Table 1: South Africa Ports of interest parameters.

Port	Location (Lon,Lat)	Area of Interest (min/max lon,lat)
Durban	(31.0316,-29.8832)	(30.5000,-29.4516,31.5854,-30.2354)
Cape Town	(18.4462,-33.9185)	(17.5131,-33.5314,18.6545,-34.1873)
Richards Bay	(32.0556,-28.8082)	(31.5656,-28.4588,32.6624,-29.2001)
Gqeberha	(25.6431,-33.9628)	(25.1228,-33.5523,26.1936,-34.3232)

3. METHODOLOGY

The study aims to explore the potential of open data to monitor ocean ships using cloud and artificial intelligence analytics. To do this, we analyse the availability of open data for vessel monitoring in strategic areas of interest.

The methodology follows several important steps. First, we define the areas of interest that we are studying. The second step is to collect open data from local and cloud sources within the study areas. For each data source, we preprocess the data on the cloud platform, clean up the datasets, and converting them to a standard format. Once all datasets are ready, we proceed to perform processing and analysis tasks on the cloud platform and evaluate the results using a combination of local and cloud resources, see illustration in Figure (3). By following this methodology, we aim to leverage cloud geoprocessing and open data sources' capabilities to improve ocean vessel monitoring efficiency and effectiveness. The following sections describe and implement each step in detail and highlight potential advantages and challenges associated with this approach.

3.1 Open-Data Collection

Numerous cloud-based platforms are available for geospatial processing and analysis. These platforms offer the advantage

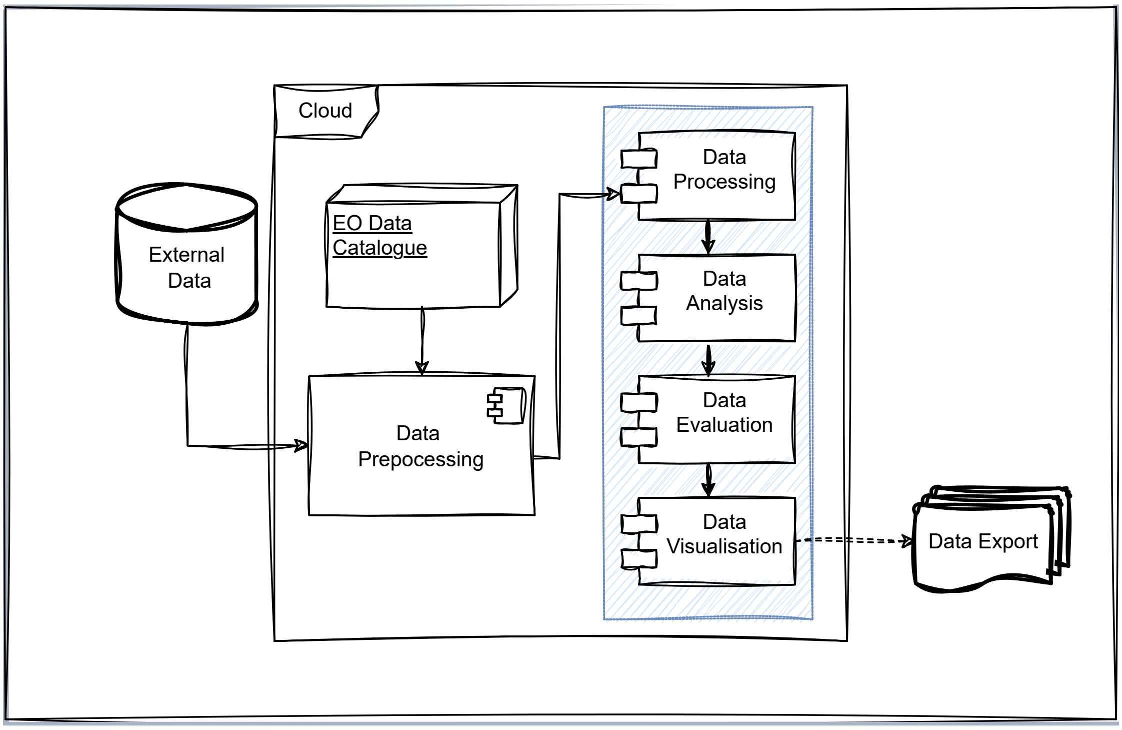

Figure 3: Cloud Environment Architecture

Figure 4: Sentinel-2 Natural Colour with Cloud Mask (orange) image.

Figure 5: A closer look of Sentinel-2 Natural Colour with ships visible at the Cape Town Port.

of storing remote sensing or spatial big data at the server level, eliminating the need for extensive data downloads. Google Earth Engine, Amazon Web Services, Microsoft Azure, Ali-Cloud, and Sentinel Hub are among the leading cloud platforms that specialise in Earth observation data (Loukili et al., 2022). In this study, we used Google Earth Engine (GEE) remote sensing data collection platform.

The GEE data catalogue consists of a large collection of publicly accessible geospatial datasets that include observations from various satellite and aerial imaging systems. In addition, users access and analyse private data using the Earth Engine API. The study used the following subsequent satellite remote sensing data:

3.1.1 SAR Data Synthetic Aperture Radar (SAR) imagery can penetrate cloud cover and provide all-weather vessel monitoring capabilities. The Sentinel-1 mission, operated by the European Space Agency (ESA), provides radar satellite data that can be used for vessel detection.

3.1.2 Optical Data Optical data enable the identification of ships through visual attributes. Despite its dependence on clear or favourable weather conditions, its high spatial resolution and ability to differentiate vessel characteristics make it invaluable for vessel detection and monitoring applications, see Sentinel-2 data in Figures (4 and 7). The Sentinel-2 mission, operated by the European Space Agency (ESA), provides high-resolution optical satellite data suitable for vessel detection.

3.1.3 Other Datasets The Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) sensor on the NASA (partnered with NOAA) satellite captures low-light imaging data, allowing the detection of illuminated vessels at sea. This capability improves vessel detection and monitoring, particularly during nighttime operations, by analysing emitted light signatures. The DNB data contribute to improved situational awareness, surveillance, and maritime safety.

3.2 Setting up cloud environment

In this study, we used Google Colab to analyse data collected from Google Earth Engine (GEE) remote sensing data collection platform(Bisong, 2019). Geemap, a Python package, was for interactive mapping with Google Earth Engine (GEE), and to convert existing GEE JavaScripts to Python scripts(Wu, 2020).

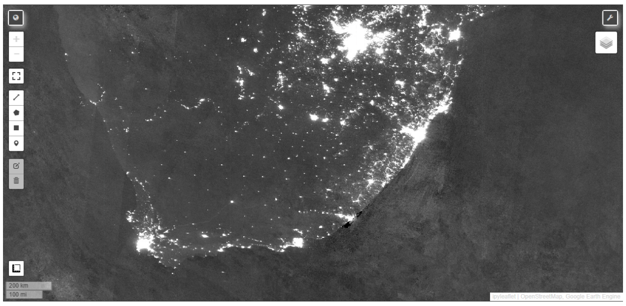

Figure 6: Monthly average radiance composite image using nighttime data from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) for January 2023

The Colaboratory, or "Colab" for short, is a product developed by Google Research. Colab allows users to write and execute Python code using the Jupyter notebook service through a browser, making it especially well-suited for machine learning, data analysis, and education. Colab provides free access to computing resources, including GPUs. (Bisong, 2019). To begin using Google Colab, follow these steps:

1. Go to the colab.research.google.com website and open a new notebook. Ensure that you are logged in with your Google account.
2. In the first code cell of your notebook, import the necessary libraries for your analysis. The commonly used libraries for Google Earth Engine (GEE) data analysis include:
   - "ee" for the Google Earth Engine Python API.
   - "folium" for visualisations.
   - "geemap" for the Google Earth Engine Python package, which provides interactive mapping capabilities.
3. Authenticate your Google Earth Engine account and initialise the Earth Engine Python API by executing the code snippet provided in the algorithm.

These steps will enable you to configure your environment and access the required libraries and resources to work with Google Earth Engine in Google Colab. The snippet of the code is illustrated in Algorithm (1).

```
# Import Libraries
import ee
import folium
import geemap

# Trigger the authentication flow.
ee.Authenticate()

# Initialize the Earth Engine API.
ee.Initialize()
```

Algorithm 1: Jupter notebook code example.

3.3 Preprocessing

The preprocessing of remote sensing satellite data for vessel detection typically involves several steps to improve the quality of the imagery and improve the accuracy of the detection algorithm. To prepare satellite imagery for vessel detection algorithms, several preprocessing steps are applied to the datasets. These steps aim to improve vessel detection by improving image quality and ensuring that all images have a standard format. The pre-processing workflow for each dataset is as follows:

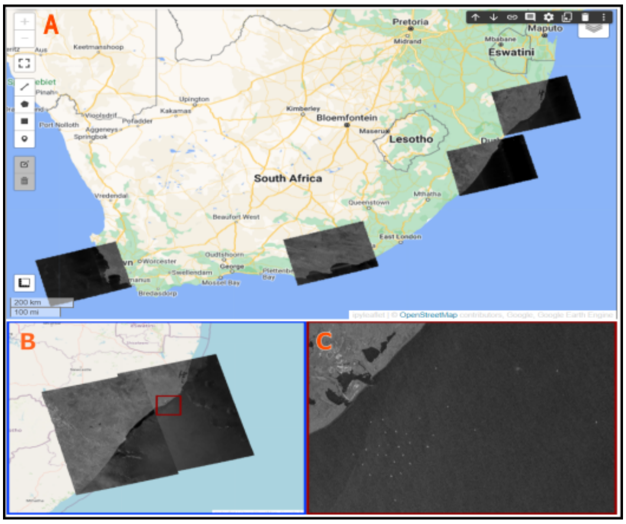

Figure 7: A) SAR imagery coverage for the selected ports of interest in South Africa. B) Mosaic SAR image coverage of the Durban and Richards Bay Ports. C) A closer look at ships (bright spots) near Richards Bay Port.

- Preprocessing of optical data includes correcting sensor-specific effects by radiometric calibration and removing distortion by geometric correction to align images with real-world coordinates. Enhancing visibility can be achieved by improving image quality, and data normalisation ensures consistency across bands.
- For SAR data, radiometric calibration addresses system-specific parameters, speckle filtering reduces noise while preserving vessel details, and geometric correction compensates for topographic variations.
- For VIIRS DNB data, radiometric calibration converts raw values and refines spatial alignment, taking into account atmospheric influences with geometric and atmospheric corrections. Stray light correction minimises contamination, increases visibility, and reduces noise, thus contributing to improved image quality.

3.4 Processing and Analysis

AI-based vessel detection algorithms use artificial intelligence techniques to autonomously detect and identify ships across various data sources, including satellite imagery, radar data, and the Automatic Identification System (AIS) data. These algorithms are purposefully designed to handle various types of data and provide comprehensive information about the presence and activity of ships. Through data analysis using artificial intelligence methods, these algorithms can identify unique vessel patterns, characteristics, and anomalies, thereby enabling efficient and accurate vessel detection. Using the cloud-based Google Colab, we were able to access GPU computing, which improves AI-based analytics, see resource usage in Figure (8

To visualize and analyse the results of the we:

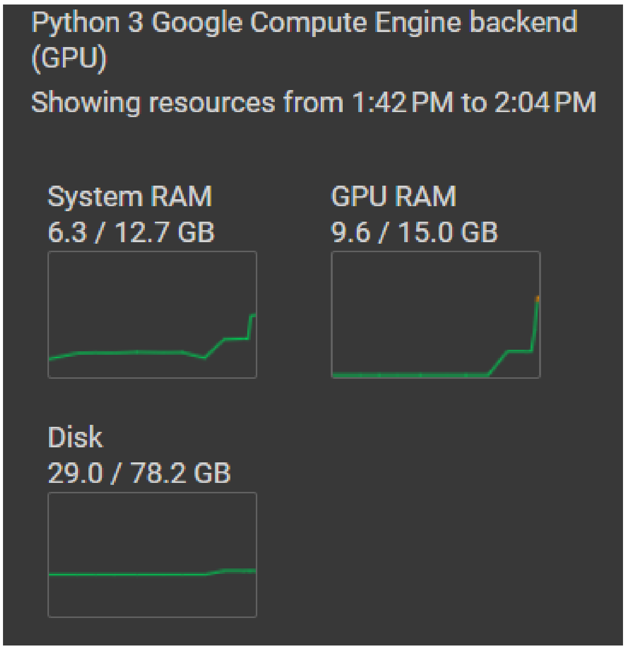

Figure 8: Python 3 Google Compute Engine backend (GPU) Showing resource usage

- Utilise tools like geemap on a cloud platform for visualising the detected vessels on an interactive map.
- Generate plots and use different visualisation parameters to analyse the results.
- Overlay the detected vessels on a basemap to gain insight into their distribution and patterns.

Following these steps, the performance of the vessel detection algorithm can be evaluated. Furthermore, visualising and analysing the results using tools such as geemap provided a comprehensive understanding of the detected vessels and assisted to identify areas for further improvement. This process ensures the accurate evaluation of the algorithm's performance, validation of the detected vessels, and continuous enhancement of the algorithm's capabilities.

4. RESULTS AND DISCUSSION

In order to investigate the potential of open-data for ocean vessel monitoring with AI analytics, the study established the following objectives:

1. Explore the technical requirements necessary for effective vessel tracking and comprehensively review the available sensors that can meet these requirements.
2. Emphasis is placed on the importance of cloud geoprocessing and the integration of spatial data access and AI analysis, aimed at improving decision-making processes and operational efficiency.
3. Provide valuable information on how ocean monitoring can significantly improve maritime safety, security, and environmental protection measures using open remote sensing data using AI analytics.

The use of AI-based vessel detection algorithms is a major advancement in maritime surveillance, fisheries management, and environmental monitoring. This technology offers essential information to improve maritime safety and security. We used the Google Earth Engine (GEE) remote sensing data collection platform to acquire geospatial data). Furthermore, we investigated the possibility of using Google Colab to analyse the GEE data obtained.

Our experiment used Google Colab to demonstrate the power of cloud technology for geospatial datasets. This platform highlighted the advantages of cloud integration, including scalability, enhanced accessibility, multiple storage options, and strong data integration features. The Colab cloud platform, combined with the Google Earth API, allowed for the smooth combination of various geospatial datasets, such as Optical, SAR, and Day/Night Band data, from sources like ESA and NASA. Accessing data is the first step to enable a complete analysis and gain meaningful insights. This gives users the opportunity to make the most of their geospatial datasets.

Cloud-based geoprocessing platforms have helped make data more accessible, but they may not be as effective as dedicated software tools such as SNAP (Sentinel Application Platform) when it comes to processing ESA Sentinel data. These specialised tools are designed to perform specific data processing and analysis tasks. To bridge the gap and improve the capabilities of geoprocessing platforms, more work needs to be done.

5. CONCLUSION

The tracking of ships in the oceans is essential for a variety of purposes, such as maritime safety, environmental monitoring, and fisheries management. This article examines the potential of open data combined with cloud-based geoprocessing and artificial intelligence (AI) analytics to track vessel activity in our oceans. We explore the advantages and difficulties of this approach and its implications for maritime monitoring and management. The study was conducted to gain insight into the potential benefits and possibilities of using open data and AI analytics for comprehensive ocean vessel monitoring. Cloud-based analysis makes it easier to access high-performance computing resources to process large geospatial data sets. The cloud not only serves as a repository for vast datasets but also a platform for efficient computation and analytics. Moreover, the integration of AI-driven vessel detection algorithms within this framework increases analytical capabilities, allowing for real-time insights and predictive modelling. By combining cloud technology and AI tools, researchers and practitioners can effectively address complex challenges in maritime surveillance, fisheries management, and environmental monitoring, leading to safer and more sustainable maritime activities.

Cloud-based geoprocessing platforms, such as Colab, offer substantial advantages in terms of data accessibility, storage, and processing capabilities. However, there is a need for data analysis tools and libraries for decision support, especially AI-based models for maritime applications. Future work includes expanding the data sets and investigating specialised geoprocessing software, such as SNAP Tools, and how they can be integrated for cloud-based analysis.

REFERENCES

Bhat, S. A., Huang, N.-F., 2021. Big data and ai revolution in precision agriculture: Survey and challenges. *IEEE Access*, 9, 110209–110222.

Bisong, E., 2019. *Google Colaboratory*. Apress, Berkeley, CA, 59–64.

Bo, L., Xiaoyang, X., Xingxing, W., Wenting, T., 2021. Ship detection and classification from optical remote sensing images: A survey. *Chinese Journal of Aeronautics*, 34(3), 145–163.

Brett, A., Leape, J., Abbott, M., Sakaguchi, H., Cao, L., Chand, K., Golbuu, Y., Martin, T. J., Mayorga, J., Myksvoll, M. S., 2020. Ocean data need a sea change to help navigate the warming world. *Nature*, 582(7811), 181–183.

Hoeser, T., Bachofer, F., Kuenzer, C., 2020. Object detection and image segmentation with deep learning on Earth observation data: A review—Part II: Applications. *Remote Sensing*, 12(18), 3053.

Huang, Q., 2020. Spatial Cloud Computing. *The Geographic Information Science & Technology Body of Knowledge*, 2nd Quarter 2020 Edition.

Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B. et al., 2022. FathomNet: A global image database for enabling artificial intelligence in the ocean. *Scientific reports*, 12(1), 15914.

Klein, T., Nilsson, M., Persson, A., Håkansson, B., 2017. From open data to open analyses—New opportunities for environmental applications? *Environments*, 4(2), 32.

Kristiansen, S., 2013. *Maritime transportation: safety management and risk analysis*. Routledge.

Li, J., Xu, C., Su, H., Gao, L., Wang, T., 2022. Deep learning for SAR ship detection: Past, present and future. *Remote Sensing*, 14(11), 2712.

Loukili, Y., Lakhrissi, Y., Ali, S. E. B., 2022. Geospatial Big Data Platforms: A Comprehensive Review. *KN-Journal of Cartography and Geographic Information*, 72(4), 293–308.

Ndehedehe, C., 2022. Cloud-based geospatial analysis. *Satellite Remote Sensing of Terrestrial Hydrology*, Springer, 73–95.

Nordling, L., 2017. South Africa tackles crime at sea with ship-spotting satellites. *Nature*.

Struwig, M., Van den Berg, A., Hadi, N., 2023. Challenges in the ocean economy of South Africa. *Development Southern Africa*, 1–15.

Vreÿ, F., Nel, M., Fouché, H., Blaine, M., 2021. South Africa: Maritime Security Sector Reform. *Capacity Building for Maritime Security: The Western Indian Ocean Experience*, 97–127.

Wang, Y., Rajesh, G., Mercilin Raajini, X., Kritika, N., Kavinkumar, A., Shah, S. B. H., 2021. Machine learning-based ship detection and tracking using satellite images for maritime surveillance. *Journal of Ambient Intelligence and Smart Environments*, 13(5), 361–371.

Wu, Q., 2020. geemap: A Python package for interactive mapping with Google Earth Engine. *Journal of Open Source Software*, 5(51), 2305.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K. X., 2019. How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, 39(6), 755–773.